JOURNAL OF VIROLOGY, Oct. 1992, p. 5714–5725
0022-538X/92/105714-12$02.00/0

Vol. 66, No. 10

Phylogenetic Analysis of 48 Papillomavirus Types and 28 Subtypes and Variants: a Showcase for the Molecular Evolution of DNA Viruses

SHIH-YEN CHAN,[1] HANS-ULRICH BERNARD,[1]* CHI-KEONG ONG,[1] SHIH-PING CHAN,[2] BIRGIT HOFMANN,[3] AND HAJO DELIUS[3]

Laboratory of Papillomavirus Biology, Institute of Molecular and Cell Biology,[1] and Department of Mathematics,[2] National University of Singapore, Singapore 0511, Singapore, and Forschungsschwerpunkt Angewandte Tumorvirologie, Deutsches Krebsforschungszentrum, Heidelberg, Germany[3]

Received 22 April 1992/Accepted 23 June 1992

Papillomaviruses are attractive models for studying the molecular evolution of DNA viruses because of the large number of isolates that exhibit genomic diversity and host species and tissue specificity. To examine their relationship, we selected two amino acid sequences, one of 52 residues within the early gene E1 and the other of 44 residues within the late gene L1, which allowed insertion- and deletion-free alignment of all accessible papillomavirus sequences. We constructed phylogenetic trees from the amino acid and corresponding nucleotide sequences from 28 published and 20 newly determined animal and human papillomavirus (HPV) genomic sequences by using distance matrix, maximum-likelihood, and parsimony methods. The trees agreed in all important topological aspects. One major branch with two clearly separated clusters contained 11 HPV types associated with epidermodysplasia verruciformis. A second major branch had all the papillomaviruses involved in genital neoplasia and, in distant relationship, the cutaneous papillomaviruses HPV type 2a (HPV-2a), HPV-3, and HPV-10 as well as the "butcher's" papillomavirus HPV-7 and two simian papillomaviruses. Four artiodactyl (even-toed hoofed mammal) papillomaviruses, the cottontail rabbit papillomavirus, and avian (chaffinch) papillomavirus type 1 formed a third major branch. Last, four papillomaviruses exhibited little affinity to any of these three branches; these were the cutaneous types HPV-1a, HPV-4, and HPV-41 and B-group bovine papillomavirus type 4. The phylogeny suggests that some branches of papillomavirus evolution are restricted to particular target tissues and that a general process of long-term papillomavirus-host coevolution has occurred. This latter hypothesis is still conjectural because of bias in the current data base for human types and the paucity of animal papillomavirus sequences. The comparison of evolutionary distances for the most closely related types with those of 28 subtypes and variants of HPV-2, HPV-5, HPV-6, HPV-16, and HPV-18 supports the type as a natural taxonomic unit, with subtypes and variants being expressions of minor intratype genomic diversity similar to that found in the natural populations of all biological species. An exception to this seems to be HPV-2c, which has an evolutionary distance from HPV-2a of the intertype magnitude and may eventually have to be regarded as a distinct type. We describe an experimental approach that estimates the taxonomic and phylogenetic positions of newly identified papillomaviruses without viral isolation and complete genomic sequencing. Finally, the paper also explores concepts for a natural, i.e., phylogenetically based, papillomavirus taxonomy; demonstrates the rooting of an HPV variant cluster against its closest type relative; and presents a hypothetical combined phylogeny of the papillomaviruses with other DNA tumor viruses.

Phylogenetic studies of viruses became increasingly feasible over the last decade with the availability of large amounts of molecular sequence information, powerful computer hardware, and sophisticated algorithms to analyze this information on the basis of specific assumptions about the process of molecular evolution (17, 19, 45, 65). The result was significant progress toward understanding the phylogeny of RNA viruses (5, 14, 15, 24, 36, 40, 58, 73) and hepadnaviruses (47). In contrast, the study of DNA virus phylogenies is still hampered by the lack of extensive DNA sequence data and the absence of a common conserved viral molecule such as the RNA-dependent RNA polymerase or the reverse transcriptase. Despite this, early attempts were made to understand relationships within the *Adenoviridae*; these attempts were based mostly on restriction enzyme cleavage patterns (55, 70). Now, for one group of DNA viruses, the papillomaviruses, the situation has perceptibly improved.

* Corresponding author.

Ever since the involvement of papillomaviruses in animal and human cancers became known (59, 65, 74), much attention has been focussed on the group. The papillomavirus sequence data base has been steadily growing subsequent to the first presentation of a phylogenetic tree based on the conserved E1 domains of nine papillomaviruses in 1986 (23). This large amount of data has made it increasingly clear that the papillomaviruses strike a nice balance between genomic conservativeness and variety. This makes phylogenetic analysis possible and the results of this analysis interesting.

By genomic conservativeness we mean that all papillomaviruses have similar genome sizes, organization, open reading frames (ORFs), and protein functions. This is of importance, as we must compare only orthologous and not paralogous or analogous sequences, because only then may we justifiably deduce the species phylogeny from what is essentially a gene phylogeny (45). It would be even better to compare several independent gene phylogenies, as we have attempted. The genomic conservativeness of papillomavi-

ruses is also the strongest evidence for their monophyletic origin. Of all the ORFs, E1 and L1 seem to be the most highly conserved of the early- and late-region genes (4, 22).

Regarding genomic variety, many (and probably most) mammal and bird species are infected by papillomaviruses (65), and most viral isolates suggest strict specificity for the host species (49). If newly discovered papillomavirus types continue to be host specific and if each host species is infected by multiple papillomavirus types (just like humans and bovines), the total number of papillomavirus types may well exceed the number of warm-blooded vertebrate species by 1 or 2 orders of magnitude, i.e., may reach 10^6 different types. Given the relative ease of identifying and sampling papillomavirus lesions and sequencing the genomes, the information reservoir for phylogenetic and taxonomic studies could be enormous.

In addition to the heterogeneity due to host species specificity, heterogeneity, i.e., papillomaviruses with significant biological differences, has been found within individual host species. For example, bovine papillomaviruses have been divided into A and B groups, the former being able to infect epithelial cells and fibroblasts and the latter restricted to epithelial cells (33, 61). Human papillomaviruses (HPVs) can be subdivided according to a tendency to infect either mucosal or cutaneous epithelia. These can be further characterized by pathologic criteria that describe the resulting neoplasia or by the differing propensities of the benign lesions to progress to malignancy (13, 50).

The plethora of papillomavirus types (in excess of 60 human types alone) demands some systematic analysis and organization (13). High on the list of aims of modern taxonomic schools is to classify "naturally" (9, 57, 62). Two of these schools, cladistics and evolutionary systematics, emphasize inherited similarities, as this is what "natural" means to them. A prerequisite for establishing a taxonomy along this line is an accurate knowledge of the organism's phylogeny. Before the advent of molecular data and methods, viral phylogenetics was impractical because viral structural details were not easily documented or interpretable in terms of identity by descent and because viruses do not leave a paleontological record. Therefore, viral taxonomy has been largely a phenetic classification based on diverse features such as morphology, host range, pathology, tissue tropism, immunological cross-reactions, etc. (44).

With the coming of age of molecular phylogenetics, the construction of a phylogenetically based viral taxonomy could begin. Analysis of the RNA viruses (5, 36, 40, 58) and the hepadnaviruses (47) brought the gratifying result that most taxonomic families of viruses, and even surmised affinities between these families, are phylogenetic relationships, in spite of the limited information that had formed the foundations of classical viral taxonomy.

Unlike most viruses, papillomaviruses have been typed on the basis of DNA hybridization (13). The criterion for a new type designation is maximally 50% homology to any of the previously known papillomavirus types in liquid hybridization experiments under stringent conditions (8). This criterion is essentially an operational one because of the lack of any controls of known sequence divergence. Lest one wonder, this is actually a very strict criterion. If the same were applied to the human adenoviruses, serotypes 1 through 31 would be amalgamated into only five types (25). It should be stressed that homology values obtained by this method are not directly comparable to similarity at the nucleotide sequence level. Subsequent sequencing has shown that papillomaviruses are highly conserved in many genomic segments (much more than what the cross-hybridization percentage might imply) but that these regions are scattered across the entire genome. Nevertheless, homology as determined by hybridization experiments has permitted assignment of each HPV type to 1 of 20 groups. Interestingly, many of these groupings show a correlation with respect to the diseases brought about by their respective members (50).

Our work sought to address several aspects of papillomavirus phylogeny. First, we asked whether alignments of different and functionally independent segments of papillomaviruses would lead to similar phylogenies. Second, we hoped that the relative positions of the papillomavirus types would generate hypotheses and suggest likely paths of papillomavirus evolution, e.g., in relationship to host speciation, tissue specificity, and pathology. Third, we hoped to investigate the type concept in the context of papillomaviruses in particular and DNA viruses in general. In this regard, it seemed desirable to compare the molecular differences among types, subtypes, and the large number of recently observed sequence variants of HPV type 16 (HPV-16) and HPV-18 (7, 10, 31, 32). Finally, a practical means of incorporating novel unsequenced papillomavirus types into the existing phylogeny was sought.

MATERIALS AND METHODS

Sequences. Table 1 groups the papillomavirus types, subtypes, and variants analyzed on the basis of some of their biomedical characteristics and independent of their genetic relatedness. Sequences for HPV-5b (72), HPV-35 (42), HPV-42 (52), and colobus monkey papillomavirus type 1 (CgPV-1) (53) were obtained from publications. The sequences for HPV-3, HPV-4, HPV-7, HPV-9, HPV-10, HPV-12, HPV-14, HPV-15, HPV-17, HPV-19, HPV-25, HPV-30, HPV-32, HPV-34, HPV-40, HPV-45, HPV-49, HPV-52, HPV-53, and HPV-56 were products of a continuing, systematic sequencing project by two of us (B.H. and H.D.). Sequences of HPV-2c (21), HPV-6a (3), HPV-6ma (37), HPV-16 variants (Sb-11, Sb-15, Bb-7, Bb-9, Tb-2, Tb-5, Gb-10, and Gb-12) (7, 31), HPV-18 variants (S18-2 and T18-15), and HPV-18 containing cervical carcinoma cell lines (HeLa and C4-1) (46) were obtained by polymerase chain reaction (PCR) amplification and sequencing using the following primer pairs: HPV-2, 5′-CCTAGTGGCTCTATG GTGTCC-3′ and 5′-ATCATATTCCTCCATATGCCT-3′; HPV-6, 5′-CCGAGCGGCTCTTTGGTGTCC-3′ and 5′-AT CATACTCTTCCACATGACG-3′; HPV-16, 5′-CCTAGTG GTTCTATGGTTACC-3′ and 5′-ATCATATTCCTCCCCAT GTCG-3′; and HPV-18, 5′-CCAAGTGGCTCTATTGTTA CC-3′ and 5′-ATCATATTCCTCAACATGTCT-3′. All other papillomavirus sequences were extracted from the GenBank data base (release 69).

Amino acid sequences with homology to the retinoblastoma protein (pRb)-binding and casein kinase II motifs of adenovirus 5 E1A and simian virus 40 large T antigen (1, 51) were extracted from the National Biomedical Research Foundation-Protein Identification Resource virus sub-data base (release 30). The sequences used for the homology search were DLTCHEAGFPPSDDEDE and NLFCSEE MPSSDDEAT from adenovirus type 5 and simian virus 40, respectively.

Phylogeny construction and evaluation. Two genomic segments, one representing part of an early and the other part of a late gene, were compared. These gene products function at different times during the viral life cycle and should have evolved by independent mechanisms. The alignment was

TABLE 1. List of 48 papillomavirus types and 28 subtypes and variants grouped by host species, site of infection, and type of lesion[a]

Host species, site of infection, and type of lesions	Papillomavirus type	Reference
Human		
Skin, EV	HPV-5, -8, -9, -12, -14, -15, -17, -19, -25, -47, -49	12, 13, 50
Skin, wart	HPV-1a, -3, -4, -10, -41	12, 13, 50
Skin and genital mucosa, wart	HPV-2	12, 13, 50
	HPV-7 (butcher's wart)	13, 13, 50
Oral mucosa, FEH	HPV-32	12, 13, 50
Nasal mucosa, inverted papilloma	HPV-57	12, 13, 50
Laryngeal or genital mucosa, papilloma	HPV-6, -11, -34, -40, -42, -53, -58	12, 13, 50
Laryngeal, carcinoma	HPV-30	12, 13, 50
Genital mucosa, preneoplasia and carcinoma	HPV-16, -18, -31, -33, -35, -39, -45, -51, -52, -56	12, 13, 42, 50
	ME180 (cell line)	
Primate: genital mucosa, papilloma	CgPV, RPV	53, 65
Artiodactyl: skin, fibropapilloma	BPV-1, BPV-2, DPV, EPV	33, 61, 65
Bovine: alimentary mucosa, papilloma	BPV-4	33, 61, 65
Other: skin, papilloma	CRPV, FPV	65
HPV-16		
Geographic variants		
Singapore	Sb-2, -5, -7, -10, -11, -13, -15, -17	7
Brazil	Bb-2, -7, -9	7
Tanzania	Tb-1, -2, -4, -5, -13, -16	7
Germany	Prototype, Gb-10 and -12	7
Cervical carcinoma cell line	CaSki	7
HPV-18		
Geographic variants		
Singapore	S18-2	46
Brazil	Prototype	46
Tanzania	T18-15	46
Cervical carcinoma cell lines	C4-1, HeLa	46
HPV subtypes	HPV-2a[b], -c	13, 21; this paper
	HPV-5a[b], -b	13, 72
	HPV-6a,-b[b], ma	13, 37

[a] Groupings are to be compared with the phylogenetic groupings of Fig. 3, 4, and 5. Because of the exigencies of space, not all primary references are given; these may be obtained by consulting reviews (13, 65) and the GenBank data base. Cell lines C4-1 and HeLa contain 1 and ~10,000 copies, respectively, of the HPV-18 genome. Abbreviations: FEH, focal epithelial hyperplasia; RPV, rhesus monkey papillomavirus.

done at the amino acid level, and the segments should not have required deletions or insertions, since such changes would add an element of ambiguity. We also favored certain highly conserved genomic regions that would permit future expansion of the analysis through partial sequences of unsequenced papillomavirus types obtained either by consensus or heterologous priming and PCR amplification (41, 56). Finally, a wide spectrum of phylogenetic methods was applied in order to avoid biasing our conclusions with respect to any given model. Extensive reviews on the methods and their relative merits are to be found in references 16, 18, 39, 45, and 66.

The ORFs examined were E7 (an oncogene) and several segments within E1 (encoding a replication protein) and L1 (encoding the major capsid protein). The detailed analysis was performed on a selected E1 and L1 segment. For a particular ORF, multiple sequence alignments were done by using the CLUSTAL V package (29), which ran on a 80486/33 MHz personal computer. Phylogenies on aligned nucleotide and amino acid sequences were then constructed by using programs in PHYLIP 3.4 (Phylogeny Inference Package) (18). This was compiled for DECsystem 5000/5500 hosts running under the ULTRIX operating system. Considerable time was saved by running separate analyses simultaneously in parallel on up to eight independent central processors.

The following phylogenetic methods were implemented. Because of the number and general similarity of the topologies, not all are presented. Specific topologies are available on request.

(i) Distance matrix. The distance matrix family of methods enjoys the advantages of short computational times (unlike maximum likelihood) and making bootstrapping of reasonable numbers and estimation of branch length in terms of nucleotide substitutions per site (unlike parsimony but similar to maximum likelihood) possible. Recent advances have been in the determination of confidence intervals for these branch lengths (unlike parsimony but similar to maximum likelihood). The method's disadvantage is the indirect use of sequence data by having to first transform nucleotide differences into a measure of evolutionary distance (unlike parsimony and maximum likelihood). We constructed neighbor-joining (54) and Fitch-Margoliash (20) trees on nucleotide sequences by using the Kimura two-parameter distance estimate (34). Ten different sequence entry orders were tried, with no difference in the topologies. The significance of subgroups of viruses within the phylogeny was assessed by bootstrap resampling (18) of 100 replicates on the complete and partial data sets (SEQBOOT, DNADIST, NEIGH BOR, FITCH, and CONSENSE programs). The bootstrap percentage quoted for a group in the figure legends is the number of trees of 100 in which that group appeared. It is meant to give a quantitative estimate of the certainty of that grouping. Bootstrapping of amino acid sequences and construction of neighbor-joining trees were executed by the PHYLOGENETIC TREES program of CLUSTAL V.

(ii) Maximum parsimony. The maximum-parsimony method constructs trees on the basis of the minimum number of nucleotide changes required to account for the data. Within practical computational limits, this often results in the generation of tens, hundreds, and even thousands of equally most-parsimonious trees, making it difficult to justify the choice of a particular tree. This difficulty was highlighted recently in the controversy over the African origin of the human maternal mitochondrial DNA lineage (2, 22, 28, 68, 69). When we did a bootstrap analysis of 100 replicates of the amino acid sequences (SEQBOOT, PROTPARS, and CONSENSE programs) and statistically evaluated it against the maximum-likelihood and neighbor-joining trees (U option in PROTPARS), the topologies were not significantly different. We also constructed a strict amino acid consensus parsimony tree from five different amino acid sequence entry orders in the manner of Hedges et al. (28).

(iii) Maximum likelihood. Maximum-likelihood trees (DNAML program) of nucleotide sequences were constructed for the whole and partial data sets. These trees were statistically evaluated by the method of Kishino and Hasegawa (35) against the neighbor-joining, Fitch-Margoliash, and maximum-parsimony trees (U option in DNAML). The topologies were not significantly different.

(iv) Method of invariants (evolutionary parsimony). Lake's linear (38) and Cavender and Felsenstein's quadratic (6) invariants (DNAINVAR) were computed for selected subsets of four sequences to determine statistical support for particular topologies under certain models of DNA sequence evolution.

Nucleotide sequence accession numbers. The complete genomic sequences of these viruses will be published elsewhere (12). The unpublished nucleotide sequences used in this study may be obtained from H.D. and were assigned the following GenBank accession numbers: 132-bp L1 segments of HPV-30 (M96279), HPV-10 (M96280), HPV-3 (M96281), HPV-12 (M96282), HPV-14 (M96283), HPV-15 (M96284), HPV-16Tb-2 (M96285), HPV-17 (M96286), HPV-18T18-15 (M96287), HPV-19 (M96288), HPV-25 (M96289), HPV-2c (M96290), HPV-32 (M96291), HPV-34 (M96292), HPV-40 (M96293), HPV-45 (M96294), HPV-49 (M96295), HPV-4 (M96296), HPV-52 (M96297), HPV-53 (M96298), HPV-56 (M96299), HPV-7 (M96300), and HPV-9 (M96301); 156-bp E1 segments of HPV-10 (M96302), HPV-12 (M96303), HPV-30 (M96304), HPV-32 (M96305), HPV-34 (M96306), HPV-3 (M96307), HPV-40 (M96308), HPV-45 (M96309), HPV-49 (M96310), HPV-4 (M96311), HPV-52 (M96312), HPV-53 (M96313), HPV-7 (M96314), HPV-9 (M96315), HPV-56 (M96316), HPV-14 (M96317), HPV-15 (M96318), HPV-17 (M96319), HPV-19 (M96320), and HPV-25 (M96321).

RESULTS

Parsimony analysis of selected papillomaviruses. As a prelude to the detailed study, we wanted to know whether functionally unrelated and nonoverlapping genomic segments contained sufficient sequence information to permit a phylogenetic analysis and if the results were consistent with relationships inferred by other experimental approaches such as DNA hybridization (49). Therefore, we constructed maximum-parsimony trees on amino acid sequences from segments of the E1, E7, and L1 genes of several selected papillomavirus types that were believed to be phylogenetically distinct. These included the closely related bovine papillomavirus types 1 and 2 (BPV-1, BPV-2) of the BPV A group; BPV-4 of the BPV B group; the deer papillomavirus (DPV); the cottontail rabbit papillomavirus (CRPV); HPV-1a, which is associated with plantar warts; HPV-5 and HPV-8, which are associated with epidermodysplasia verruciformis (EV); HPV-6 and HPV-11, which are associated with condylomata acuminata; and HPV-16 and HPV-18, which are associated with cervical intraepithelial neoplasia and cervical cancer.

All phylogenetic trees (data not shown) had consistently the same topologies. All animal papillomaviruses were on one branch. The mucosal papillomaviruses HPV-6, HPV-11, HPV-16, and HPV-18 were on another branch, with HPV-6 and HPV-11 closer to each other than either was to HPV-16 or HPV-18. HPV-5 and HPV-8 were on a third major branch, with HPV-1a on a branch remote from all others. We concluded that a detailed analysis of these genomic segments would be worthwhile and would give generally consistent and representative results.

Alignment of conserved polypeptide segments of the E1 and L1 genes of 48 papillomavirus types and 2 subtypes. Within the viral early region, the transforming genes E5, E6, and E7 have been found to be quite variable. In contrast, large conserved sequence blocks exist in genes E1 and E2, probably because of the steric requirements of the replication and transcription functions (23, 27). Figure 1 shows the alignment of a 52-amino-acid segment from the E1 gene. This particular segment was selected because it could be aligned without postulating any insertions or deletions. This was also the only sequence available in this genomic region for CgPV-1 (53). We regarded the inclusion of this monkey virus important (one of only two non-human-primate viruses). Unfortunately, E1 sequence information was unavailable for the avian (chaffinch) papillomavirus FPV-1, which, however, was represented in the L1 alignment.

The L1 gene codes for the major capsid protein. Different papillomaviruses share a common antigen that is delimited by the C-terminal half of this molecule (64). L1 sequences are so well conserved that four different primer pairs could be found for amplifying known and unknown papillomavirus types (56, 63). We decided to base the phylogenetic analyses of L1 nucleotide sequences on the amino acid sequence alignment shown in Fig. 2; the FPV-1 sequence, however, requires the postulate of a deletion. CgPV-1 sequences were not available for this L1 region.

Features of papillomavirus phylogeny. Figures 3, 4, and 5 show distance matrix, maximum-likelihood, and consensus parsimony trees. There are three major groups. One contains all nonprimate animal papillomaviruses (BPV-1, BPV-2, DPV, elk papillomavirus (EPV), and CRPV [the animal group]). The second contains most of the HPV types associated with EV (HPV-5a, HPV-5b, HPV-8, HPV-12, HPV-14, HPV-19, HPV-25, and HPV-47 [EV group]). Four other EV-associated HPVs (HPV-9, HPV-15, HPV-17, and HPV-49) form a separate cluster at the base of the EV group branch in Fig. 3 and 5 or at the base of the animal group branch in Fig. 4. Figure 4 and all the L1 trees suggest that the EV group may be further subdivided into two subgroups, one containing HPV-5, HPV-8, HPV-12, and HPV-47 and the other containing HPV-14, HPV-19, and HPV-25. The third major group contains 28 HPVs and monkey papillomaviruses mainly associated with genital mucosal lesions (the mucosal group). Within this group but often on a separate branch are HPV-2a, HPV-2c, HPV-3, and HPV-10, which cause cutaneous lesions (see Discussion). In separate bootstrap analyses on data sets of all combinations of the three major groups, taking two at a time, we found the mucosal

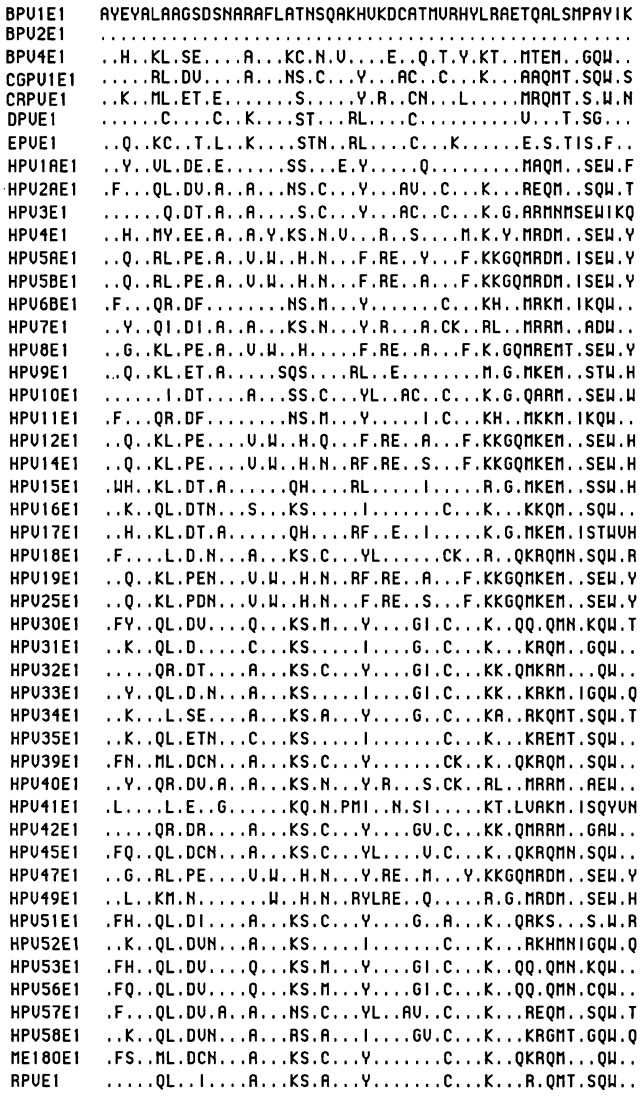

FIG. 1. Multiple sequence alignment of a 52-amino-acid segment of the E1 ORF used to construct the E1 phylogenies. In HPV-16, this segment corresponds to genomic positions 1989 to 2114. No HPV-2c or FPV sequences were available for this region. ME180 is an integrated HPV genome in the ME180 cervical-carcinoma cell line. Abbreviations used are elaborated in Table 1, footnote *a*.

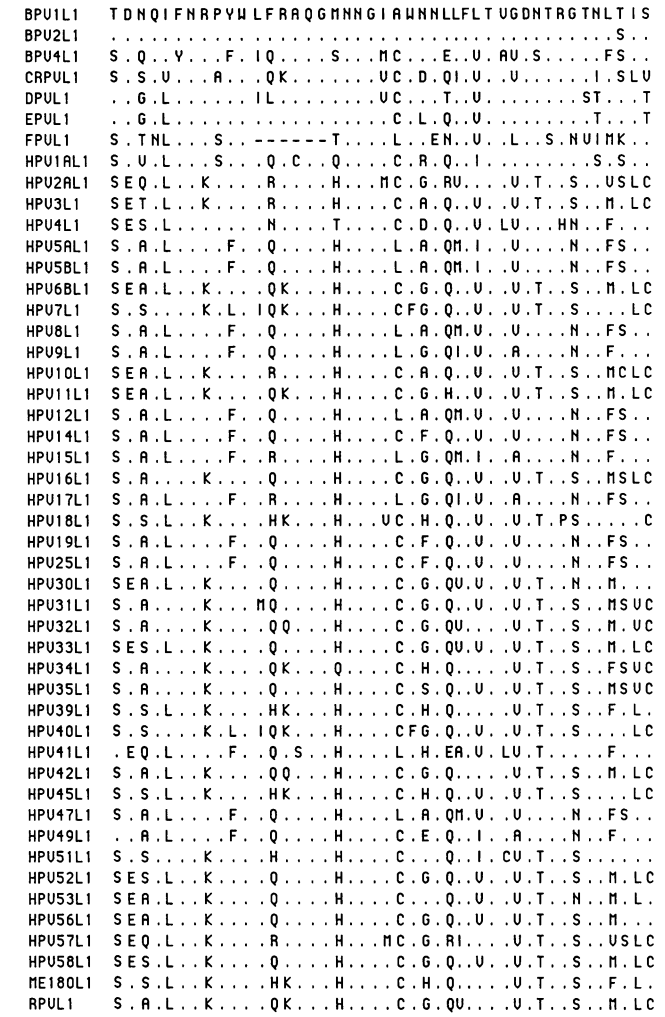

FIG. 2. Multiple sequence alignment of a 44-amino-acid segment of the L1 ORF used to construct the L1 phylogenies. Abbreviations are as in Table 1. No CgPV sequences were available for this region. In HPV-16, this region corresponds to genomic positions 6540 to 6671.

group separated from the animal group at the 97% level and from the EV group at the 93% level. The EV and animal groups were separated at the 89% level.

With respect to the general features described above, all L1 and E1 trees were very similar. The two obvious differences were the placement of HPV-51 within the mucosal group (E1 distance matrix trees) and of CRPV between the EV and mucosal groups and distantly related to HPV-4 (L1 maximum-likelihood tree only).

Within the mucosal group, we consistently found small two- and three-member groups. Among these were HPV-2a, HPV-2c, and HPV-57; HPV-3 and HPV-10; HPV-6b and HPV-11; HPV-7 and HPV-40; HPV-32 and HPV-42; HPV-30, HPV-53, and HPV-56; HPV-33, HPV-52, and HPV-58; HPV-16, HPV-31, and HPV-35; and HPV-18, HPV-39, HPV-45, and ME180. Bootstrapping and simple inspection of numerous trees revealed that there was no consistent pattern to the positions of these small groups with respect to each other. However, in all trees, CgPV-1, HPV-2a, HPV-57, HPV-3, and HPV-10 were grouped together. All E1 trees grouped HPV-16, HPV-31, HPV-35, HPV-33, HPV-52, and HPV-58 together. Rhesus monkey papillomavirus was placed near the base of the HPV-2a–HPV-3–CgPV-1 lineage in the L1 trees. The relative positions of the other groups were so variable as to defy simple description.

The animal group is less closely knit because of host species differences and fewer samples. The consistent groups were BPV-1–BPV-2, DPV, and EPV. This artiodactyl cluster was also significantly separated from BPV-4 and the two monkey papillomaviruses in the mucosal group.

In addition, a comparison of many E1 and L1 trees show that there was a heterogenous group of cutaneous papillomaviruses (BPV-4, HPV-1a, HPV-4, and HPV-41) whose positions were not consistently defined and depended on whether DNA or amino acid data were used, which ORF was examined, and which method of tree construction was

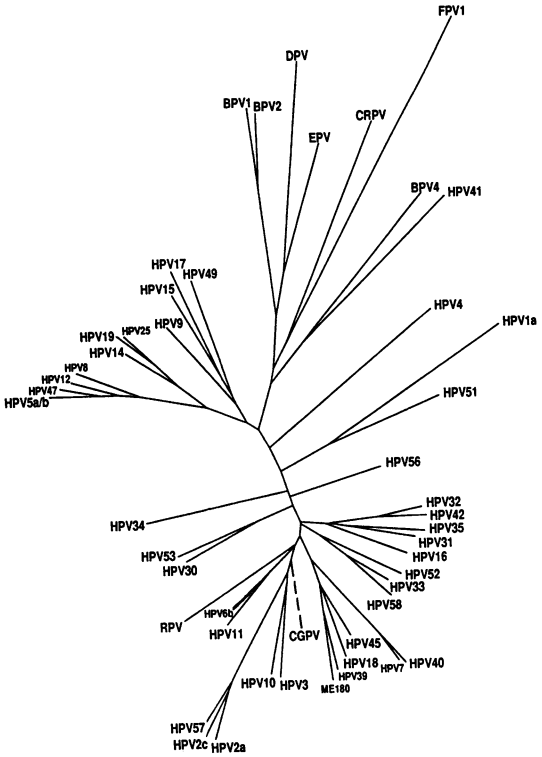

FIG. 3. Neighbor-joining distance matrix tree based on nucleotide sequences corresponding to the L1 amino acid alignment of Fig. 2. Ten different sequence orders were tried; no differences in topologies were found. Branches are freely rotatable about their internodes without any effect on topology. Because of constraints in the tree-plotting program, relative branch lengths are only approximately correct; exact lengths are available on request. The dotted line is the position of CgPV1 as estimated from the E1 tree. The E1 tree was similar in all major respects. Bootstrap percentages are given for the following groups: HPV-5, -8, -12, and -47 (95%); HPV-14, -19, and -25 (80%); BPV-1, BPV-2, DPV and EPV (94%); HPV-2 and -57 (99%); HPV-3 and -10 (100%); HPV-2, -3, -10, and -57 (75%); HPV-6b and -11 (68%); HPV-7 and -40 (100%); HPV-32 and -42 (88%); HPV-30, -53, -56 (98%); HPV-16, -31, and -35 (81%); HPV-18, -39, and -45 and ME180 (81%); and HPV-33, -52, and -58 (75%). Of the 132 nucleotide sites, a closely related pair such as HPV-19 and HPV-25 differed at 12 sites, a moderately distant pair such as HPV-16 and HPV-18 differed at 32 sites, and a distantly related pair such as CRPV and DPV differed at 43 sites. RPV, rhesus monkey papillomavirus.

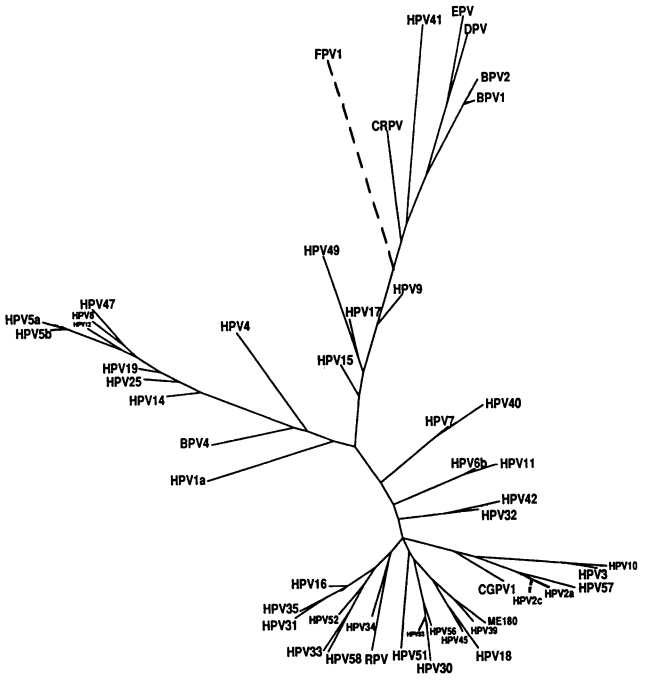

FIG. 4. Maximum-likelihood tree based on nucleotide sequence corresponding to the E1 amino acid alignment of Fig. 1. Differences between this tree and that in Fig. 3 are primarily a result of the different methods of construction and are discussed in the text. This tree was not statistically different from the distance matrix and parsimony trees. The dotted lines are the positions of FPV-1 and HPV-2c as estimated from the L1 tree. Of the 156 nucleotide sites, a closely related pair such as HPV-19 and HPV-25 differed at 21 sites, a moderately distant pair such as HPV-16 and HPV-18 differed at 38 sites, and a distantly related pair such as CRPV and DPV differed at 66 sites. RPV, rhesus monkey papillomavirus.

used. Generally, they occupied positions between the animal-EV groups and the mucosal group. BPV-4 was closely associated with HPV-41 and HPV-4 in the L1 and E1 trees, respectively. It could be found in the animal or EV groups, between them, or between the animal-EV groups and the mucosal group. HPV-1a was placed in the animal group or between the animal and EV groups in most trees.

Incorporating novel papillomaviruses into the phylogeny with consensus primers. We wanted to determine the feasibility of using consensus priming and PCR to determine the phylogenetic position of novel papillomaviruses without completely sequencing them. The primer pair PCR2 was chosen because it seemed to readily pick up unknown HPV types (56). This primer pair amplifies a DNA segment that codes at its 5′ end for the 29 C-terminal amino acids of the 44-amino-acid alignment of Fig. 2. We constructed distance matrix, parsimony, and maximum-likelihood trees from the 87 nucleotides that coded for this amino acid segment and found that these trees exhibited the same general and specific features previously described (data not shown).

During an ongoing study of the intratype diversity of HPV-16 and HPV-18 (7, 46), we received clones tentatively identified as HPV-18 by hybridization. After sequencing the long control region (LCR), we found no similarity to HPV-18; nevertheless, consensus priming with PCR2 and sequencing of the L1 indicated a very close phylogenetic relationship to HPV-18. Subsequently, we received known HPV-45 sequences which matched that of our unknown clone. We conclude that the PCR2 primer pair is a useful tool for checking whether a suspected new papillomavirus is in fact novel and for obtaining preliminary information about its relative position in the papillomavirus phylogeny.

Distinction between types, subtypes, and variants. Evolutionary distance is a measure of divergence between two phylogenetically linked organisms. It is usually calculated from the proportion of nucleotide or amino acid differences between the two (66). Various corrections are made to allow for multiple substitutions at the same site and various rates of substitution between different nucleotide pairs (39).

To compare the evolutionary distances between HPV types, subtypes, and variants, we amplified by PCR, cloned, and sequenced the 132-nucleotide segment that corresponded to the amino acid segment shown in Fig. 2 for HPV-2c (21); HPV-6a (3); HPV-6ma (37); eight HPV-16 variants that represented the African, German, and East

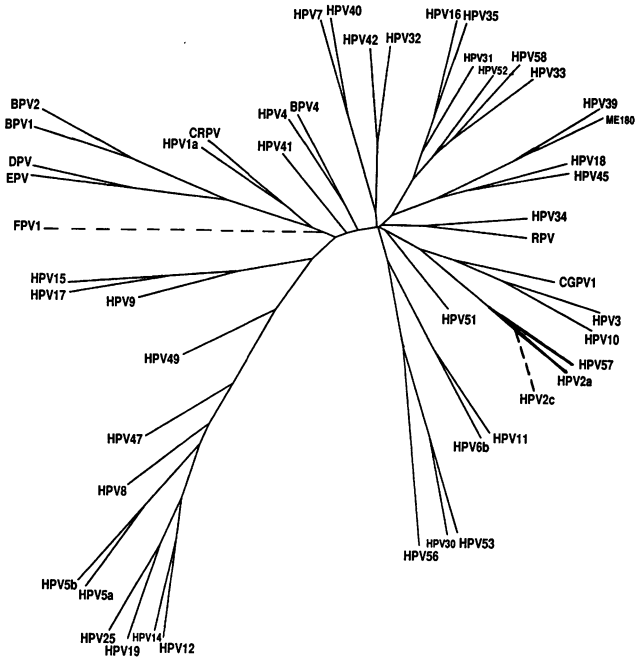

FIG. 5. Consensus maximum-parsimony tree based on 100 replicates of the amino acid sequence of the E1 segment given in Fig. 1. As this is a consensus tree, the branch lengths have no significance; only the branching order (groupings) can be interpreted. Dotted lines are the positions of FPV-1 and HPV-2c as estimated from the L1 tree. Bootstrap percentages are given for the following groups: HPV-5, -8, -12, -14, -19, -25, and -47 (85%); HPV-9, -15, and -17 (72%); BPV-1, BPV-2, DPV, and EPV (87%); HPV-7 and -40 (100%); HPV-32 and -42 (68%); HPV-6b and -11 (99%); HPV-2a and -57 (87%); HPV-30, -53, and -56 (87%); and HPV-18, -39, and -45 and ME180 (83%). RPV, rhesus monkey papillomavirus.

```
HPV16       TCTGATGCCCAAATATTCAATAAACCTTATTGGTTACAACGAGCACAGGG
HPV16var    ..................T.....................................
HPV35       ..C.....A........T........A........G.....T.....A..
HPV18       TCTGACTCCCAGTTGTTTAATAAACCATATTGGTTACATAAGGCACAGGG
HPV18var    .....................................................
HPV45       .....T..T..A..A........G.....................C.....
HPV2a       TCTGAACAACAGTTGTTTAATAAGCCTTACTGGCTACGGAGGGCCCAAGG
HPV2c       ........G..A..............C.....................G..
HPV57       ........G.....A.....C.............G...........G..
HPV5a/b     AGTGATGCTCAATTGTTTAACAGGCCCTTCTGGCTCCAAAGAGCCCAAGG
HPV47       ............C.............A..............G..T..G..
HPV6a/b/ma  TCTGAGGCACAATTGTTTAATAAGCCATATTGGCTACAAAAAGCCCAGGG
HPV11       ..A.....T.....A........A...........T.....G..T.....
BPV1        ACTGATAATCAAATTTTTAATCGGCCCTACTGGCTATTCCGTGCCCAGGG
BPV2        ..............A.....CA....T..T.........A.G..T.....

HPV16       CCACAATAATGGCATTTGTTGGGGTAACCAACTATTTGTTACTGTTGTTG
HPV16var    .....................................................
HPV35       ...T........T.........A........T.G...........A....
HPV18       TCATAACAATGGTGTTTGCTGGCATAATCAATTATTTGTTACTGTGGTAG
HPV18var    ............A.C..................................
HPV45       C............A....T...........G..G...........A..G.
HPV2a       GCACAACAATGGTATGTGCTGGGGCAATAGGGTCTTTCTGACTGTGGTGG
HPV2c       A..T..T.....................C..A.................
HPV57       A..T........C..............C..A....C..A..A.......
HPV5a/b     TCATAATAATGGCATCCTGTGGGCTAATCAAATGTTTATCACAGTGGTTG
HPV47       ..............T.....................G.......T..A.
HPV6a/b/ma  ACATAACAATGGTATTTGTTGGGGTAATCAACTGTTTGTTACTGTGGTAG
HPV11       ................C.....A..C..C...................
BPV1        CATGAACAATGGAATTGCATGGAATAATTTATTGTTTTTAACAGTGGGGG
BPV2        ...........G.....................A........T..A....

HPV16       ATACTACACGCAGTACAAATATGTCATTATGT
HPV16var    ................................
HPV35       ....A..C..T.............TG.G...
HPV18       ATACCACTCCCAGTACCAATTTAACAATATGT
HPV18var    ................................
HPV45       .C..T..C.G......T.........T.....
HPV2a       ACACCACACGTAGCACTAATGTATCTCTGTGT
HPV2c       .....CA...G..T..C.....C.........
HPV57       .......G..C.....A.....C...T.....
HPV5a/b     ACAACACAAGAAATACTAATTTCAGTATTTCT
HPV47       ................A........C..C...
HPV6a/b/ma  ATACCACACGCAGTACCAACATGACATTATGT
HPV11       ................A..T......C.....
BPV1        ACAATACACGTGGTACTAATCTTACCATAAGT
BPV2        .T..C.....G..A.....C....GT..T...
```

FIG. 6. Multiple sequence alignment of the 132-nucleotide L1 segment, corresponding to the sequence in Fig. 1, of selected papillomavirus types subtypes and variants. Only differences are specified; periods indicate no change with respect to the sequence above.

Asian branches of the HPV-16 phylogeny (7, 31); two HPV-18 variants; and two HPV-18 cervical-carcinoma cell lines (46). Figure 6 is the L1 sequence alignment, and Table 2 compares the evolutionary distances between the closest and most significant pairs of papillomaviruses. The closest related types differ by 9.8 to 17% substitutions per site. Subtypes and variants are closer by 1 order of magnitude.

The HPV-16 and HPV-18 variants chosen for the comparison were those that showed a maximal divergence (up to 5%) in their LCR (the transcriptional regulatory region) sequences (7, 31, 46), a clear example of the greater variability in noncoding regions than in protein-coding regions. Strangely, the HPV-2a–HPV-2c subtype comparison showed an evolutionary distance of the intertype magnitude. HPV-2c has been previously described as an isolate significantly different from HPV-2a (21), and our analysis suggests that it could be provisionally regarded as a distinct type.

We have shown that presently described subtypes and variants are quantitatively identical expressions of intratype diversity and that the decision on whether an ambiguous isolate represents a new type or a subtype can be further assessed simply and quantitatively.

Rooting of the HPV-16 variant cluster against HPV-31. In the simplest models, molecular evolution is believed to occur by the gradual accumulation of single mutations. For two closely related papillomavirus types, e.g., HPV-16 and HPV-31, this model seems to predict a continuum of genomic variants extending from the time of the most recent common ancestor to the present day. The intratype variability that we observed may be partial evidence of this genomic continuum. Consequently, it might be possible to identify in a collection of variants those that are closest to their common ancestor with another type.

We had previously reported (7) that an analysis of the

TABLE 2. Evolutionary distances between the most significant and the closest pairs of papillomaviruses[a]

Sequences compared	Distance (%)[b]
HPV-16 vs HPV-35	16.3
HPV-18 vs HPV-45	15.9
HPV-2a vs HPV-57	15.1
HPV-5a and -b vs. HPV-47	9.8
HPV-6 vs HPV-11	12.5
BPV-1 vs BPV-2	17.4
HPV-2a vs HPV-2c	12.4
HPV-16 variants	0.8
HPV-18 variants	2.3
HPV-5a vs HPV-5b	0
HPV-6a vs HPV-6b and -6ma	0

[a] The L1 segment shown in Fig. 2 was used.

[b] The Kimura two-parameter distance estimates are given as percent substitutions per site and were calculated by the DNADIST program of PHYLIP 3.4. Maximum-likelihood distances were almost identical, and distances calculated for the E1 segment were comparable except for BPV-1 and BPV-2, for which the sequences have not diverged so extensively (7.4%), and for HPV-5a and HPV-5b, for which the sequences have diverged more (6.8%) than in other subtypes or variants.

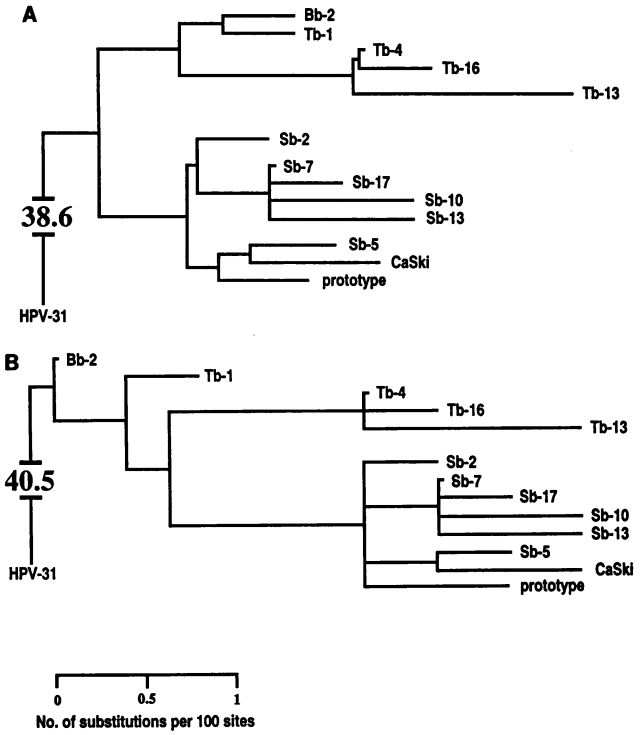

FIG. 7. Phylogeny of the HPV-16 E5 variant tree with HPV-31 E5 nucleotide sequences as the out-groups calculated by Fitch-Margoliash distance matrix method (A) and maximum-likelihood method (B). Only the horizontal lengths indicate distance. The distance to the the out-group is not to scale and is given between the horizontal bars.

LCR and E5 ORF supported the separation of HPV-16 variants (according to LCR sequences) into deep African and Eurasian branches. In order to generate hypotheses on possible roots of the HPV-16 variant tree, we used the E5 coding sequence from the phylogenetically closest type, HPV-31, to root distance matrix and maximum-likelihood trees of HPV-16 E5 variants (Figure 7; for sequence data, see Fig. 6 and 7 in reference 7). We could not use the LCR, E1, or L1 sequence to root the variant tree, because the HPV-31 LCR is too divergent and the E1 and L1 sequences of the HPV-16 variants are too conserved. The root is indicated where the out-group (HPV-31) joins the variant tree. In Fig. 7A, this is between the African and Eurasian HPV-16 variants; in Fig. 7B, this is among the African variants. The most likely interpretation of this is that the common ancestor of the group is none of the presently identified HPV-16 variants (e.g., certainly not the prototype) but is a yet-unidentified and possibly extinct genome at or close to the root. Figure 7 also illustrates the great difference between intertype and intratype evolutionary distances. It supports the taxonomic status of the type because there is no present-day detectable genomic continuum between closely related types, possibly because of extinction of the intermediates.

Phylogenetic relations among papillomaviruses and other DNA tumor viruses. Molecular evidence to support the common origin of many separate families of RNA viruses has been gradually accumulating (5, 14, 36). In contrast, no such relationship among the DNA virus families or even among genera within a family, e.g., papillomaviruses and polyomaviruses, is obvious. This general lack of molecular similarity and common genomic organization is remarkable, since DNA viruses are generally observed to evolve more slowly than RNA viruses by several orders of magnitude (60). It could imply several independent origins for DNA viruses, a very ancient common origin with subsequent loss of information about common ancestry due to diversification and specialization, or some combination of these two scenarios. It seems reasonable to speculate that phenetically similar genera, especially those currently grouped together in a single family, have a common ancestry. Within such a group, one would expect to find sequence homologies in domains of viral proteins with similar functions. Recently, just such a similarity was discovered within proteins which bind the retinoblastoma gene product (pRb) from three different groups of DNA tumor viruses (papillomavirus, simian virus 40 polyomavirus, and adenovirus) and a human fibroblast cell line (1, 11, 51). We did a 100-replicate bootstrap protein parsimony analysis (Fig. 8B) based largely on the alignment described in reference 11 and given in Fig. 8A. This topology should be viewed strictly as a hypothesis-generating device only, as the sequence similarities within these large-T-antigen, E1a, and E7 proteins could have resulted from convergent evolution, and some of these motifs have yet to be shown to bind pRb. However, we propose that the topology is not an unreasonable initial hypothesis, as it groups the HPVs in the same associations as they had been described by traditional taxonomic criteria. This has been confirmed by using a much larger data set of DNA tumor virus sequences (data not shown). Using this larger data set, we see the adenoviruses and also various polyomaviruses grouped together. The human sequences might correspond to a distant out-group if we hypothesize that these viral sequences had a cellular origin. We note that many of the EV-group HPVs have very poor sequence similarity in this region and that the group A BPVs have no discernible similarity at all.

DISCUSSION

The impressive host species diversity of papillomaviruses could be viewed as evidence for a general pattern of long-term papillomavirus-host coevolution. We ourselves had previously published (7) evidence to support an ancient association between HPV-16 and humans. If one accepts concordance in the topologies of phylogenies derived from parasites and their hosts as evidence for coevolution, then the best evidence comes from the cutaneous papillomaviruses of the artiodactyls, BPV-1, BPV-2, DPV, and EPV. Superficially, the positions of FPV-1, CRPV, and the primate papillomaviruses might also be taken as support. However, we caution against overinterpretation for several reasons. (i) We have too few data to accurately place the animal papillomaviruses. Hidden diversities are likely to be revealed by future sampling for animal papillomaviruses related to the HPV mucosal group and EV groups and more human non-EV cutaneous viruses. The BPV A and B groups and the HPVs are evidence for this diversity already. (ii) The lagomorphs and birds are represented by only one member each. (iii) The simian papillomaviruses are not distinguished from the HPVs, and only two members have been analyzed. (iv) Finally, the human cutaneous papillomaviruses (HPV-1a, HPV-4, and HPV-41) are not securely placed phylogenetically and may indicate the existence of additional major branches which may also contain yet-unidentified animal papillomaviruses.

A

HPV-5a	DLFCEEE-LPN--E-QET-EEE
HPV-8	DLLCEEE-LPN--E-QET-EEE
HPV-6b	GLHCYEQ-LVD------SSEDE
HPV-11	GLHCYEQ-LED------SSEDE
HPV-16	DLYCYEQ-LND------SSEEE
HPV-18	DLLCHEQ-LSD-------SEEE
HPV35	DLYCYEQ-LCD------SSEEE
BPV-4	NLHCEEE--------IET-EEV
CRPV	SLHCDEA-LENL------SDDD
Ad4	DLRCYEECLPP-------SDDE
Ad5	DLTCHEAGFPP-------SDDE
Ad40	DLKCYEDGLPP-------SDPE
SV40	NLFCSEE-MP-------SSDDE
SLPV	DLFCSET-MS------SSSDED
PBK	DLFCHED-MFA-------SDEE
PJC	DLFCHEE-MFA-------SDDE
RBP-1	-LVCHEVDLDDLDEKDKTSIED
RBP-2	NLFCDEE-IP-----IK-SEEV

B

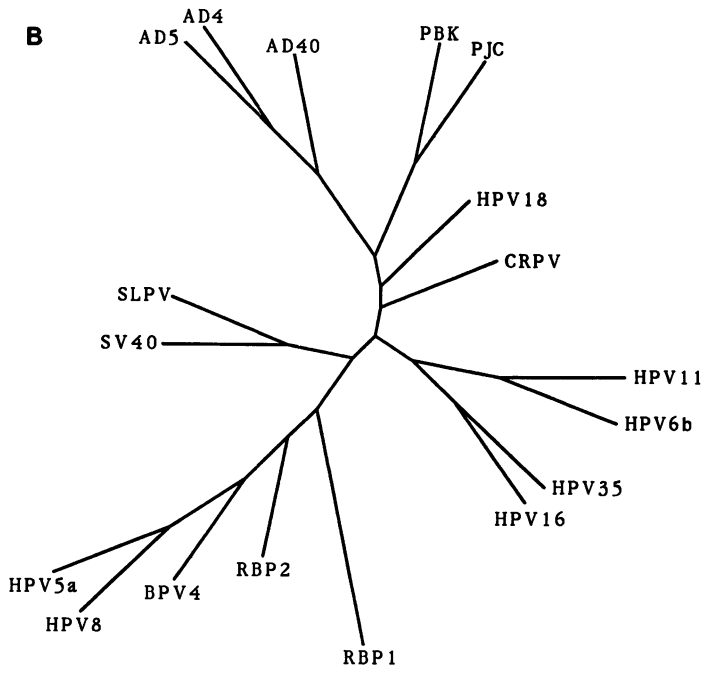

FIG. 8. (A) The multiple alignment of the pRb-binding and casein kinase II motifs found in proteins of some DNA tumor viruses and human lung fibroblasts is essentially that given in reference 11. This corresponds to conserved region 2 of the adenovirus E1A proteins. Note that the casein kinase II motif of retinoblastoma binding protein type 2 as indicated in reference 11 is 5′ to the pRb-binding motif and therefore could not be aligned; instead, a similar 3′ sequence was chosen. (B) Consensus maximum-parsimony tree of 100 replicates based on this alignment. Abbreviations: Ad, adenovirus; PBK, polyoma virus BK; PJC, polyomavirus JC; RBP, retinoblastoma binding protein from human lung fibroblasts; SLPV, simian lymphotropic virus), SV40 (simian virus 40).

In a few cases, the diversification of papillomaviruses into different host species may have been initiated by lateral transfer. The fact that BPV-4 shares a most recent common ancestor with HPV-4 or HPV-41 in some phylogenies and that the simian papillomaviruses are closely related to some mucosal HPVs might be indicative of this. Viable interhost species transfer does seem to occur regularly for influenza virus (43) and as rare and relatively recent events for the human and simian immunodeficiency viruses (57) and for some protozoans of the *Plasmodium* genus (71). In the case of papillomaviruses, evidence (49) indicates that lateral transfer normally does not lead to productive infections.

Interpreting the phylogenetic diversity of papillomaviruses vis à vis host species diversity is further complicated by an added layer of diversity within a single host species. This is evident when papillomaviruses of a single host species are extensively studied, as in the cases of HPV and BPV. In fact, the outstanding feature of HPV phylogeny is the evolutionary separation of papillomaviruses along lines of tissue tropism and pathology, namely, the major EV, cutaneous, and mucosal groupings. The distinctiveness of the last group is further supported by the presence of a genomic feature which is functionally independent of E1 and L1: a characteristic alignment of binding sites for the cellular transcription factor SP1 and two dimeric viral E2 proteins at the E6 promoter. This appears to be a strongly conserved regulatory motif (67).

Our analysis leads us to view the EV-associated HPVs as presently subdivided into two subgroups. The smaller group contains HPV-9, HPV-15, HPV-17, and HPV-49. The larger one may be further subdivided into two, one containing HPV-5, HPV-8, HPV-12, and HPV-47 and the other containing HPV-14, HPV-19, and HPV-25. This subdivision has been previously recognized in the analysis of LCR sequences (16). In contrast, the mucosal group is poorly resolved at all but the tips of its branches. These represent several previously recognized affinities, e.g., HPV-2 and -57; HPV-6 and -11; HPV-16, -31, and -35; and HPV-18, -39, and -45 (13, 42, 50) and some new ones, e.g., HPV-3 and -10; HPV-7 and -40; HPV-32 and -42; HPV-33, -52, and -58; and HPV-30, -53, and -56. No group consistently defined along lines of tissue specificity (cutaneous versus mucosal epithelia) or disease prognosis (low or high propensity for malignant conversion) exists. This conclusion is also supported by our distance matrix and parsimony analyses of the E7 ORF (data not shown) and a separate parsimony analysis of the E6 and E7 genes (42). Indeed, the picture is not inconsistent with the various subgroups being representatives of at least four different lineages that have diverged from each other within a short time at an early stage (an explosion or star phylogeny). Included in the mucosal group are some consistently placed cutaneous viruses (HPV-2a, HPV-2c, HPV-3, and HPV-10). In all our trees, HPV-2a, which also occurs in oral mucosal lesions (13, 30), has always been closest to HPV-57, a mucosal HPV. This consistency suggests that the placement is not artifactual and that the mucosal-cutaneous

barrier may have been breached a number of times in the past. This genetic relatedness is mirrored by their pathological manifestations. HPV-2a and HPV-57 cause verruca vulgaris (exophytic common warts), and HPV-3 and HPV-10 cause the very characteristic verruca plana (papular flat warts).

HPV-7 is a cutaneous-mucosal HPV (13, 26) closely associated with the mucosal group. The association of HPV-7 infection with the handling of animal material, e.g., in butchers, gave rise to speculation that this occupational group was infected by an unidentified animal papillomavirus that could occasionally cross species barriers, given extensive exposure. The consistent position of HPV-7 near or in the mucosal group does not support this speculation and suggests that correlation of HPV-7 with professional activity arises through some mechanism other than exposure to an animal papillomavirus reservoir. More data on animal papillomaviruses, particularly porcine and bovine mucosal-group papillomaviruses, would be helpful.

That BPV-4 is evolutionarily distinct from BPV-1 and BPV-2 is consistent with its different pathology, i.e., epithelial papillomas versus fibropapillomas (33, 61). Correlation of this virus's genomic sequence with those of other animal or HPV genomes is so poor that it strongly supports the traditional subdivision of BPV types into A and B groups. Affinities of BPV-4 to HPV-4 or HPV-41 that seem to suggest relatedness should be cautiously interpreted, because the distances between them are very large. The assignments of these viruses may become clear only after sampling of more closely related genomic types.

When the phylogenetic grouping of the HPVs based on small conserved genomic segments is compared with the grouping based on total genomic hybridization (50), we find almost complete agreement. Although the hybridization approach is rapid and potentially uses all the genomic information available, the phylogenetic-sequencing approach has much higher resolution and permits one to construct a natural hierarchy of relatedness. However, one needs to select the segment(s) to be analyzed with care, and the analysis is potentially more time-consuming. In fact, the two approaches are complementary; knowing that one is dealing with a type, subtype, or particular group would be helpful in designing heterologous primers for determining the E1 or L1 sequence.

At one level, molecular evolution of genomes is a continuum of events. Genomes that arise in sequential generations accumulate sequentially occurring mutations, first A, then B, then C, etc. Eventually, the unmutated parent genome and a progeny genome with the accumulated mutations A, B, C, etc., cannot be clearly linked unless the intermediate genomes are found. It is an enigma of speciation that intermediate genomes that link presently living organisms can normally not be found, possibly because of extinctions. The result is that existing populations appear as more or less isolated genomes such as species or types (but see reference 15).

In view of the proliferation of papillomavirus types, subtypes, and variants despite stringent typing procedures and the renewed interest in the establishment of a baseline for the reclassification of the genus (13), we sought to quantify in evolutionary terms the present nature of the type, subtype, and variant distinction. According to Table 2, intratype differences in the E1 and L1 segments range from 0 to 2.3%, while the closest intertype differences range from 9.8 to 17.4%. To a certain degree, this apparent evolutionary gap between types is due to the present operational definition of a papillomavirus type. Therefore, the possibility that careful sampling will eventually lead to the discovery of a continuum between the genomic sequences of some types, such as BPV-1 and BPV-2 or HPV-6 and HPV-11, cannot be excluded. Alternatively, if the nature of papillomavirus type "speciation" is different from subtype or variant diversification, the papillomavirus type as we know it would represent a true biological entity. Research in this area could begin with an examination of isolates that cross-hybridize to the extent of 60 to 90%, e.g., HPV-10, HPV-10P, and HPV-10PW (48). We note, however, that our published (7, 31) and ongoing studies of the diversity of more than 200 HPV-16 and HPV-18 isolates showed a maximum of only 5% divergence in the LCR and even less in the L1 gene. The initial sampling was unbiased, so that HPV-16 or HPV-18 genomes divergent by as much as 10% would have been detected if there had been any.

In summary, we believe that papillomavirus types that exist today constitute natural biological taxonomic units. However, like any population, they contain genomically slightly diverse members. The reasons for the apparent absence of genomes that mediate between types are unclear, as are the mechanisms of speciation. The tremendous host species diversity and the topology of the papillomavirus phylogeny indicate an element of viral-host coevolution over long periods, possibly as long as the host's evolution itself. At least for HPVs, there is also evolution along lines of tissue tropism and pathology. Further sampling will be necessary to refine this picture.

ACKNOWLEDGMENTS

We thank Herbert Pfister for helpful discussions and the gift of HPV-2c, HPV-6a, and HPV-6ma DNA; Joseph Felsenstein for advice; the two reviewers for invaluable comments; and Pek Chong-Kuan, Richard Tan, and Seow Kah-Tong for computing assistance.

REFERENCES

1. **Barbosa, M. S., C. Edmonds, C. Fisher, J. T. Schiller, D. R. Lowy, and K. H. Vousden.** 1990. The region of the HPV E7 oncoprotein homologous to adenovirus E1a and SV40 large T antigen contains separate domains for Rb binding and casein kinase II phosphorylation. EMBO J. **9**:153–160.
2. **Barinaga, M.** 1992. "African Eve" backers beat a retreat. Science **255**:686–687.
3. **Boshart, M., and H. zur Hausen.** 1986. Human papillomaviruses in Buschke-Lowenstein tumors: physical state of the DNA and identification of a tandem duplication in the noncoding region of a human papillomavirus 6 subtype. J. Virol. **58**:963–966.
4. **Broker, T. R., and L. T. Chow.** 1987. Human papillomaviruses of the genital mucosa: electron microscopic analyses of DNA heteroduplexes formed with HPV types 6, 11 and 18, p. 589–594. *In* B. M. Steinberg, J. L. Brandsma, and L. B. Taichman (ed.), Cancer cells 4: DNA tumor viruses. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
5. **Bruenn, J. A.** 1991. Relationships among the positive strand and double-strand RNA viruses as viewed through their RNA-dependent RNA polymerases. Nucleic Acids Res. **19**:217–226.
6. **Cavender, J. A., and J. Felsenstein.** 1987. Invariants of phylogenies in a simple case with discrete states. J. Classif. **4**:57–71.
7. **Chan, S.-Y., L. Ho, C.-K. Ong, V. Chow, B. Drescher, M. Dürst, J. ter Meulen, L. Villa, J. Luande, H. N. Mgaya, and H. U. Bernard.** 1992. Molecular variants of human papillomavirus-16 from four continents suggest pandemic spread of the virus and its coevolution with humankind. J. Virol. **66**:2057–2066.
8. **Coggin, J. R., Jr., and H. zur Hausen.** 1979. Workshop on papillomaviruses and cancer. Cancer Res. **39**:545–546.
9. **Crowson, R. A.** 1970. Classification and biology. Heinemann Educational Books, London.
10. **Deau, M. C., M. Favre, and G. Orth.** 1991. Genetic heteroge-

neity among human papillomaviruses associated with epidermodysplasia verruciformis: evidence for multiple allelic forms of HPV-5 and HPV-8 E6 genes. Virology **184**:492–503.
11. **Defeo-Jones, D., P. S. Huang, R. E. Jones, K. M. Haskell, G. A. Vuocolo, M. G. Hanobik, H. E. Huber, and A. Oliff.** 1991. Cloning of cDNAs for cellular proteins that bind to the retinoblastoma gene product. Nature (London) **352**:251–254.
12. **Delius, H.** Unpublished data.
13. **de Villiers, E.-M.** 1989. Heterogeneity of the human papillomavirus group. J. Virol. **63**:4898–4903.
14. **Doolittle, R. F., D. F. Feng, M. S. Johnson, and M. A. McClure.** 1989. Origins and evolutionary relationships of retroviruses. Q. Rev. Biol. **64**:1–30.
15. **Dopazo, J., F. Sobrino, E. L. Palma, E. Domingo, and A. Moya.** 1988. Gene encoding capsid protein VP1 of foot-and-mouth disease virus: a quasispecies model of molecular evolution. Proc. Natl. Acad. Sci. USA **85**:6811–6815.
16. **Essner, A., and H. Pfister.** 1990. Epidermodysplasia verruciformis associated human papillomaviruses present a subgenus specific organization of the regulatory region. Nucleic Acids Res. **18**:3919–3922.
17. **Felsenstein, J.** 1982. Numerical methods for inferring evolutionary trees. Q. Rev. Biol. **57**:379–404.
18. **Felsenstein, J.** 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution **39**:783–791.
19. **Felsenstein, J.** 1988. Phylogenies from molecular sequences: inference and reliability. Annu. Rev. Genet. **22**:521–565.
20. **Fitch, W. M., and E. Margoliash.** 1967. Construction of phylogenetic trees. Science **155**:279–284.
21. **Fuchs, P. G., and H. Pfister.** 1984. Cloning and characterization of papillomavirus type 2c DNA. Intervirology **22**:177–180.
22. **Gee, H.** 1992. Statistical cloud over African Eden. Nature (London) **355**:583.
23. **Giri, I., and O. Danos.** 1986. Papillomavirus genomes: from sequence data to biological properties. Trends Genet. **2**:227–232.
24. **Gorman, O. T., W. J. Bean, Y. Kawaoka, I. Donatelli, Y. Gou, and R. G. Webster.** 1990. Evolution of the influenza A virus nucleoprotein genes: implications for the origins of H1N1 human and classical swine viruses. J. Virol. **65**:3704–3714.
25. **Green, M., J. K. Mackey, W. S. Wold, and P. Rigden.** 1979. Thirty-one human adenovirus serotypes (Ad1–Ad31) form five groups (A–E) based upon DNA genome homologies. Virology **93**:481–492.
26. **Greenspan, D., E.-M. de Villiers, J. S. Greenspan, Y. G. De Souza, and H. zur Hausen.** 1988. Unusual HPV types in oral warts in association with HIV infection. J. Oral Pathol. **17**:482–487.
27. **Ham, J., V. Dostatni, J. M. Gauthier, and M. Yaniv.** 1991. The papillomavirus E2 protein: a factor with many talents. Trends Biochem Sci. **16**:440–444.
28. **Hedges, S. B., S. Kumar, and K. Tamura.** 1992. Technical comments: human origins and analysis of mitochondrial DNA sequences. Science **255**:737–739.
29. **Higgins, D. G., A. J. Bleasby, and R. Fuchs.** 1992. CLUSTAL V: an improved software for multiple sequence alignment. Submitted for publication.
30. **Hirsch-Behnam, A., H. Delius, and E. M. de Villiers.** 1990. A comparative sequence analysis of two human papillomavirus (HPV) types 2a and 57. Virus Res. **18**:81–98.
31. **Ho, L., S.-Y. Chan, V. Chow, T. Chong, S.-K. Tay, L. L. Villa, and H. U. Bernard.** 1991. Sequence variants of human papillomavirus type 16 in clinical samples permit verification and extension of epidemiological studies and construction of a phylogenetic tree. J. Clin. Microbiol. **29**:1765–1772.
32. **Icenogle, J. P., P. Sathya, D. L. Miller, R. A. Tucker, and W. E. Rawls.** 1991. Nucleotide and amino-acid sequence variation in the L1 and E7 open reading frames of human papillomavirus type 6 and type 16. Virology **184**:101–107.
33. **Jackson, M. E., W. D. Pennie, R. E. McCaffery, K. T. Smith, J. Grindlay, and M. S. Campo.** 1991. The B subgroup bovine papillomaviruses lack an identifiable E6 open reading frame. Mol. Carcinogenesis **4**:382–387.
34. **Kimura, M.** 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol. **16**:111–120.
35. **Kishino, H., and M. Hasegawa.** 1989. Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in Hominoidea. J. Mol. Evol. **29**:170–179.
36. **Koonin, E. V.** 1991. The phylogeny of RNA-dependent RNA polymerases of positive-strand RNA viruses. J. Gen. Virol. **72**:2197–2206.
37. **Kulke, R., G. E. Gross, and H. Pfister.** 1989. Duplication of enhancer sequences in human papillomavirus 6 from condylomas of the mamilla. Virology **173**:284–290.
38. **Lake, J.** 1987. A rate-independent technique for the analysis of nucleic acid sequences: evolutionary parsimony. Mol. Biol. Evol. **4**:167–191.
39. **Li, W.-H., C.-C. Luo, and C.-I. Wu.** 1985. Evolution of DNA sequences, p. 1–84. *In* R. J. MacIntyre (ed.), Molecular evolutionary genetics. Plenum Press, New York.
40. **Mahy, B. W. J. (ed.).** 1991. Related viruses of the plant and animal kingdoms. Semin. Virol. **2**:1–77.
41. **Manos, M. M., Y. Ting, D. K. Wright, A. J. Lewis, T. R. Broker, and S. M. Wolinsky.** 1989. The use of polymerase chain reaction amplification for the detection of genital human papillomaviruses, p. 209–214. *In* M. Furth and M. Greaves (ed.), Cancer cells 7: molecular diagnostics of human cancer. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
42. **Marich, J. E., A. V. Pontsler, S. M. Rice, K. A. McGraw, and T. W. Dubensky.** 1992. The phylogenetic relationship and complete nucleotide sequence of human papillomavirus type 35. Virology **186**:770–776.
43. **Murphy, B. R., and R. G. Webster.** 1990. Orthomyxoviruses, p. 1091–1154. *In* B. N. Fields and D. M. Knipe (ed.), Fields virology. Raven Press, New York.
44. **Murphy, F. A., and D. W. Kingsbury.** 1990. Virus taxonomy, p. 9–35. *In* B. N. Fields and D. M. Knipe (ed.), Fields virology. Raven Press, New York.
45. **Nei, M.** 1987. Molecular evolutionary genetics. Columbia University Press, New York.
46. **Ong, C.-K., S.-Y. Chan, J. ter Meulen, and H. U. Bernard.** Unpublished data.
47. **Orito, E., M. Mizokami, Y. Ina, E. N. Moriyama, N. Kamashima, M. Yamamoto, and T. Gojobori.** 1989. Host independent evolution and a genetic classification of the hepadnavirus family based on nucleotide sequences. Proc. Natl. Acad. Sci. USA **86**:7059–7062.
48. **Ostrow, R. S., K. R. Zachow, S. Watts, M. Bender, F. Pass, and A. J. Farras.** 1983. Characterization of two HPV-3 related papillomaviruses from common warts that are distinct clinically from flat warts of epidermodysplasia verruciformis. J. Invest. Dermatol. **80**:436–440.
49. **Pfister, H.** 1987. Papillomaviruses: general description, taxonomy, and classification, p. 1–18. *In* N. P. Salzman and P. M. Howley (ed.), The *Papovaviridae*, vol. 2. The papillomaviruses. Plenum Press, New York.
50. **Pfister, H.** 1990. Molecular biology of genital HPV infections, p. 38–49. *In* G. Gross, S. Jablonska, H. Pfister, and H. Stegner (ed.), Genital papillomavirus infections. Springer-Verlag, Berlin.
51. **Phelps, W. C., C. L. Yee, K. Munger, and P. M. Howley.** 1988. The human papillomavirus type 16 E7 gene encodes transactivation and transformation functions similar to adenovirus E1a. Cell **53**:539–547.
52. **Philipp, W., N. Honore, M. Sapp, S. T. Cole, and R. E. Streeck.** 1992. Human papillomavirus type 42: new sequences, conserved genome organisation. Virology **186**:331–334.
53. **Reszka, A. A., J. P. Sundberg, and M. E. Reichmann.** 1991. In vitro transformation and molecular characterization of colobus monkey venereal papillomavirus DNA. Virology **181**:787–792.
54. **Saitou, N., and M. Nei.** 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4**:406–425.
55. **Sambrook, J., M. Sleigh, J. A. Engler, and T. R. Broker.** 1980.

The evolution of the adenovirus genome. Ann. N.Y. Acad. Sci. **354**:426–452.

56. **Schiffman, M. H., H. M. Bauer, A. T. Lorincz, M. M. Manos, J. C. Byrne, A. G. Glass, D. M. Cadell, and P. M. Howley.** 1991. Comparison of Southern blot hybridization and polymerase chain reaction methods for the detection of human papillomavirus DNA. J. Clin. Microbiol. **29**:573–577.
57. **Scott-Ram, N. R.** 1990. Transformed cladistics, taxonomy and evolution. Cambridge University Press, Cambridge.
58. **Sharp, P. M., and W.-H. Li.** 1988. Understanding the origins of AIDS viruses. Nature (London) **336**:315.
59. **Shope, R. E., and E. W. Hurst.** 1933. Infectious papillomatosis of rabbits. J. Exp. Med. **58**:607–624.
60. **Smith, D. B., and S. C. Inglis.** 1987. The mutation rate and variability of eukaryotic viruses: an analytical review. J. Gen. Virol. **68**:2729–2740.
61. **Smith, K. T., and M. S. Campo.** 1985. The biology of papillomaviruses and their role in oncogenesis. Anti-Cancer Res. **5**:31–48.
62. **Sneath, P. H., and R. R. Sokal.** 1973. Numerical taxonomy. W. H. Freeman & Co., San Francisco.
63. **Snijders, P. J. F., C. J. L. M. Meijer, and J. M. M. Walboomers.** 1991. Degenerate primers based on highly conserved regions of amino acid sequence in papillomaviruses can be used in a generalized polymerase chain reaction to detect productive human papillomavirus infections. J. Gen. Virol. **72**:2781–2786.
64. **Strike, D. G., W. Bonnez, R. C. Rose, and R. C. Reichman.** 1989. Expression in *Escherichia coli* of seven DNA fragments comprising the complete L1 and L2 open reading frames of human papillomavirus 6b and localization of the common antigen region. J. Gen. Virol. **70**:543–555.
65. **Sundberg, J. P.** 1987. Papillomavirus infections in animals, p. 40–103. *In* K. Syrjanen, L. Gissmann, and L. G. Koss (ed.), Papillomaviruses and human disease. Springer-Verlag, Berlin.
66. **Swofford, D. L., and G. J. Olsen.** 1990. Phylogeny reconstruction, p. 411–501. *In* D. M. Hillis and C. Moritz (ed.), Molecular systematics. Sinauer Associates, Sunderland, Mass.
67. **Tan, S. H., B. Gloss, and H. U. Bernard.** 1992. During negative regulation of the human papillomavirus-16 E6 promoter, the viral E2 protein can displace Sp1 from a proximal promoter element. Nucleic Acids Res. **20**:251–256.
68. **Templeton, A. R.** 1992. Technical comments: human origins and analysis of mitochondrial DNA sequences. Science **255**:737.
69. **Vigilant, L., M. Stoneking, H. Harpending, K. Hawkes, and A. C. Wilson.** 1991. African populations and the evolution of human mitochondrial DNA. Science **253**:1503–1507.
70. **Wadell, G., M. L. Hammarskjoeld, G. Winberg, T. M. Varsanyi, and G. Sundell.** 1980. Genetic variability of adenoviruses. Ann. N.Y. Acad. Sci. **354**:16–42.
71. **Waters, A. P., D. G. Higgins, and T. F. McCutchan.** 1991. *Plasmodium falciparum* appears to have arisen as a result of lateral transfer between avian and human hosts. Proc. Natl. Acad. Sci. USA **88**:3140–3144.
72. **Yabe, Y., A. Sakai, T. Hitsumoto, H. Kato, and H. Ogura.** 1991. A subtype of human papillomavirus 5 (HPV-5b) and its subgenomic segment amplified in a carcinoma: nucleotide sequences and genomic organizations. Virology **183**:793–798.
73. **Yamashita, M., M. Krystal, W. Fitch, and P. Palese.** 1988. Influenza B virus evolution: cocirculating lineages and comparison of evolutionary pattern with those of influenza A and C viruses. Virology **163**:112–122.
74. **zur Hausen, H.** 1991. Viruses in human cancers. Science **254**: 1167–1173.